Monitoring Performance and Mental Workload in an Automated System

Indramani L. Singh, Anju L. Singh, and Proshanto K. Saha

Cognitive Science Laboratory, Department of Psychology
Banaras Hindu University, Varanasi – 221 005
ilsingh_bhu@rediffmail.com

Abstract. Human performance in computer-aided system has engrossed inevitably human issues in cognitive functioning. The present endeavor focuses on the associated influence of training, automation reliability on the monitoring performance and workload in multi-task ambience. MAT battery was utilized with engine-system monitoring, two dimensional tracking, and fuel resource management tasks were the concerned elements, in which only system engine-monitoring task was automated in the training as well as in the final test sessions. A 2 x 2 x 2 x 3, mixed factorial design was employed. Monitoring performance, false alarms, reaction time and root mean square error performance were recorded as dependent measures. Results revealed that automation-induced complacency might be the feature of multi-task condition where subjects detected automation failures under high static system reliability. Results further showed that mental workload significantly reduced from pre- to post-sessions.

Keywords: Automation, Complacency Workload, Monitoring Performance.

1 Introduction

Automation technology is a pervasive phenomenon, which has engrossed inevitably human issues of cognitive functioning. Automation envisages the thought of electronic replacement of human operator. Evidently, automation altered person's attitude towards the functioning of an automated machine, thereby placing more reliance and conviction on such systems [16]. Besides this another notion of automation has also been put forth as the execution of functions by a machine (preferably a computer) that was previously carried out by a human resource [11]. The performance in automation scenario depends on the interaction of people with the advanced technology. Moreover, the relationship between automation and mental workload of operator is an imperative consideration in respect to efficiency and safety in many modern human machine systems. For example, cockpit automation has made it possible to reduce flight times, increase fuel efficiency, navigate more effectively, and extend or improve the pilot's perceptual and cognitive capabilities [20, 21]. It is noteworthy that the benefits derived from automation use come after paying certain costs also [12], for instance, increased monitoring demands, unbalanced trust,

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 426–435, 2007.

cognitive overload, decision biases, skill degradation etc. Over trust of automation is sometimes referred to as complacency, which occurs when people trust the automation more than what is warranted and can result in very severe negative consequences, if the automation is less than fully reliable [12, 11]. Complacency can lead to a decreased monitoring of the system and a decreased likelihood of detecting system malfunctions. Moreover, introduction of automation in complex systems is embraced as reduction of the workload, and thereby reducing the human error. Instead, Woods [23] argued that automation merely changes how work is accomplished. Wiener [22] has even claimed that in some instances, the introduction of automation may increase the workload.

Bearing in mind such elements, researchers [12], put forth that any performance consequences for complacency were more likely to exist in a multi-task ambience. They examined the effect of levels of automation reliability i.e., constant and variable, on automated monitoring performance in various experimental conditions. Poorer monitoring efficiency was observed under high constant automation reliability compared to the variable. The results suggested that automation-induced complacency was more easily detectable in a multi-task environment when operators were engaged in performing numerous tasks. Considering the issue Singh et al. [18] presented automated task at the centre by spatial superimposition of monitoring task over the tracking, which could eliminate monitoring inefficiency (complacency). Results indicated that automation-induced complacency was not primarily influenced by the location, to be monitored for the automated task. Later, Singh, Molloy and Parasuraman [17], conducted another study and they showed that the centrally located monitoring task could not improve monitoring performance in automation mode, which suggested the robust nature of automation-induced complacency phenomenon.

Training is another important issue relevant to automation-induced complacency. Automation can place conflicting demand upon pilots, with which they may not be well trained to meet (e.g., passive monitoring versus active control) unless they have been specifically trained to cope with these demands. It has been suggested that inadequate training may lead to several automation-induced problems in the cockpit. For instance the negative effect of automation on monitoring performance may be related in part to a lack of 'automation based' skills. Recently, Sharma and Singh [14], examined the role of increased amount of manual training and automation reliability on a flight simulation task. Results indicated that, there was no benefit of extended manual training on automated complacency.

The rationale behind introducing automation in complex system is the reduction of workload and hence thereby reducing the human error propensity. Despite the logical and intuitive rationale that operators will choose automation under heavy workload, studies both of pilots and non-pilots performing laboratory [13], and aviation like tasks [3], revealed little, if any, tendency to choose automation more often at higher levels of task demand. It is possible that the influence of workload on automation use may emerge only when the workload is experienced for a sustained period of time. Another possibility is that more complex attributes of workload in real environments such as workload management and trade-offs need to be modeled in the laboratory in order to more fully comprehend the impact of workload on the use of automation. Thus, automated system can both reduce and increase mental workload. For instance, pragmatically it has been observed that glass cockpits in commercial aircraft have relieved workload in areas such as reduced display clutter, and enhanced automated flight procedures [7]. However, the same cockpit systems can amplify workload by presenting operators with more options in their task and causing mode uncertainty [6]. Hence, these studies revealed that the automation might/or might not affect workload.

1.1 Present Study

The underlying principle behind the present endeavor focuses on the associated influence of extended automated training, automation reliability on the monitoring performance and workload in multi-task ambience. Some studies have suggested that automation ought to be designed with the objective to reduce operator's mental workload, while some other studies have indicated that automation does not necessarily reduce workload. In view of these contentious issues about the role of training, reliability and workload on the detection of automation failures, an effort has been made to examine the effects of extended automated training and system reliability on the relationship between monitoring automation failure and mental workload. Primarily, it was hypothesized that, increased automation training would reduce automation-induced complacency; secondly, automation induced complacency would be higher in constant system reliability condition than in variable system reliability along with increase over time periods in constant system reliability condition than in variable system reliability condition and, finally, automation would reduce mental workload.

2 Methods and Procedure

Participants: Eighty non-pilots with normal (20/20) or corrected to normal visual acuity, aged 19 to 25 years, volunteered in this study. Subjects were randomly assigned in each of the four experimental conditions. Each subject received 10-min manual practice on flight simulation task, besides 3-min demo.

2.1 Flight Simulation Task

A revised version of multi-attribute task battery (MATB), [2] was used in the present study. Multi-attribute task battery is a flight simulation package, comprising engine-system monitoring, two dimensional compensatory tracking, fuel resource management, communications, and scheduling tasks. The modified version of MATB allows each component task to be performed either manually or under automation mode. In the present study, only engine-system monitoring task was automated in the training as well as in the final test sessions. These three tasks were presented in separate windows on a 14" SVGA color monitor of a PC-486 computer.

2.2 NASA-Task Load Index Scale

NASA Task Load Index (NASA-TLX), [4] was administered before and after the final test session individually for the assessment of mental workload. The reliable index of overall was .83. In TLX, workload is defined as the 'cost incurred by human operators to achieve a specific level of performance.'

2.3 Design

A 2 (Training) x 2 (Automation reliability) x 2 (Sessions) x 3 (Blocks) mixed factorial design was employed with repeated measures on the last two factors. Training (30-min and 60-min) and automation reliability (constant and variable) were treated as between subject factors with sessions (two 30-min) and blocks (six blocks, each of 10-min.) as within-subjects factors. Automation reliability was defined as the percentage of correct detection of malfunctions by the automation routine in each 10-min block in the engine-system monitoring task. The rate of automation reliability was constant (87.5%) from block to block in constant system reliability and in the variable system reliability automation varied from high (87.5%) to low (56.25%) and low to high from block to block alternately. Each subject was administered six 10-min blocks in two successive 30-min sessions.

2.4 Procedure

Out of eighty non-pilots, 40 were given short- automation training and remaining 40 were given long training. Furthermore, out of 40 non-pilots in each training group, 20 were randomly assigned the constant system reliability and the remaining 20 subjects were assigned the variable system reliability.

In the automated test session, the subjects were informed that only engine-system monitoring task would be automated and they were instructed to pay attention to only on tracking and fuel management tasks. Subjects were also told that automation routine was less than 100% reliable and in case of automation failure they had to detect the malfunctions, if any, and to reset the system-monitoring task immediately, by pressing a designated function key within 10 seconds. The correct detection (hits rate), incorrect detection (false alarms) of malfunctions, monitoring reaction time (RT) and root mean square (RMS) error were recorded as dependent measures.

3 Results

Means and standard deviations for correct detection of automation failure on all components of MATB were calculated for six 10-min automated blocks. Mean values of the correct detection of monitoring task performance showed that the subjects detected slightly more malfunctions in the long-training-variable reliability condition (M = 59.15; SD = 33.98) than in its counterparts i.e. short-training-variable reliability condition (M = 58.98; SD = 33.98). Similarly, subjects detected high number of malfunctions in the long-training-constant reliability condition (M = 42.50; SD = 37.32) than in the short-training-constant reliability condition (M = 39.24; SD = 34.27).

The mean correct detection performance further demonstrated that the subjects detected slightly more malfunctions in the long training (M = 50.83, SD = 37.29) as compared to the short-training condition (M = 49.12, SD = 35.82), irrespective of system reliability. Furthermore, subjects' detection accuracy was higher in the variable system reliability condition (M = 59.07, SD = 34.03) than in the constant reliability condition (M = 40.87, SD = 35.66), irrespective of training. These mean performances further revealed that the mean correct detection performance decreased from 3% to 15% after 30-min in the three experimental conditions except in the long-training-variable reliability condition. It is also evident from the mean performance that decrement in performance across blocks appeared from 23% to 40% in the constant reliability, whereas 4% to 5% deterioration emerged at some point of time among six blocks (see Fig. 1).

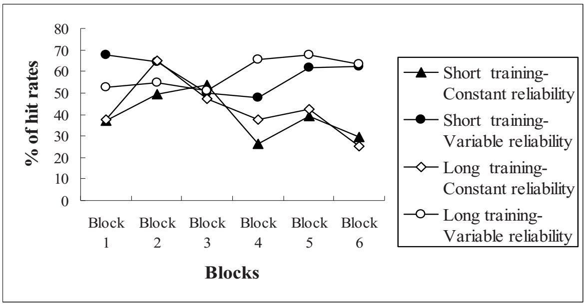

Fig. 1. Correct detection performance (hit rates) under four experimental groups during automated test sessions

Monitoring performance data were then computed for a 2 x 2 x 2 x 3 analysis of variance. The ANOVA showed that the main effect of training was not significant, $F_{(1, 76)} = .13$; ns, thereby suggesting that the amount of automation training given prior to the subjects under constant or variable automation reliability conditions has no impact on monitoring inefficiency (automation-induced complacency). Thus, the present finding does not support the first hypothesis that the high amount of automation training would reduce automation-induced complacency (see Fig. 2). Results further revealed that the main effect of system reliability was found highly significant $F_{(1, 76)} = 14.88$; $p < .01$. This result also indicated that the subjects had high reliance on automation while automation reliability was constant from block to block resulting in poor accuracy (more complacent). Contrarily, in variable reliability condition the reliance of subjects was varying from high to low and low to high, so subjects allocated more attention in detection of automation failures, resulting in better monitoring performance (less complacent). This finding supported second hypothesis, which maintained that the automation-induced complacency would be higher in the constant system reliability than it would be in the variable reliability.

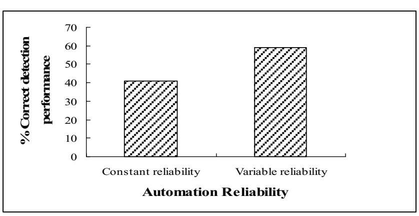

Fig. 2. Correct detection performance as a function of automation (system) reliability

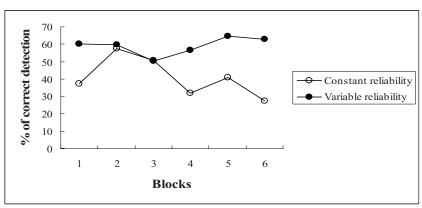

Fig. 3. Correct detection performance as a function of automation reliability across 10-min block

The interaction of system reliability and sessions, $F_{(1, 76)} = 9.39$; $p < .01$, was also found to be significant suggesting that types of system reliability i.e., constant and variable reliability impaired monitoring performance across sessions resulting in automation-induced complacency after a session of 30-min. The system reliability by session and by block interaction effect was also significant, $F_{(2, 152)} = 3.13$; $p < .05$ (see Fig. 3). The main effect of block was further found highly significant which propounded that automation-induced complacency appeared across time periods. This finding explained that automation-induced complacency increased across sessions and blocks, which supported the third hypothesis stating that automation-induced complacency would increase across time periods in constant reliability condition. The obtained findings on false alarms performance indicated that automation had no impact on performance however, the effect of system reliability suggested that the subjects committed more false alarms in the variable system reliability than in the constant reliability.

The ANOVA findings on reaction time performance showed that the main effect of training was significant, which suggested the benefit of the long training over short training. The system reliability and session interaction also reached at significance level, $F_{(1, 76)} = 4.65$; $p < .05$, which indicated that system reliability affected the speed of response in detecting correct automation failures in the flight simulation task across sessions.

The RMSE performance on tracking task suggested no benefit of training and system reliability. However, tracking performance showed an improvement across two sessions, $F_{(1, 76)} = 12.69$; $p< .01$. The ANOVA results on fuel performance indicated that the main effect of training was significant, $F_{(1, 76)} = 3.88$; $p< .05$. This finding suggested that the increased amount of training improved fuel performance. The results also indicated the benefits of either types of training over six 10-min blocks, $F_{(2, 152)} = 5.71$; $p< .01$. However, the main effect of system reliability was not significant, $F_{(1, 76)} = .003$; ns. This finding indicated no effect of system reliability on fuel resource management performance. The interaction effect of training and system reliability was also not significant, $F_{(1, 76)} = .58$; ns. This finding further showed that the length of training and system reliability could not enhance fuel performance.

3.1 Perceived Mental Workload

The NASA-TLX includes the process evaluating of relative importance of the six subscales by each subject to calculate a weighted mean. Miyake and Kumashiro [10], reported a high correlation (r = 0.971) between the weighted mean computed in the NASA-TLX and the simple arithmetic mean ratings of the scale. Their results suggested that the mean rating can be considered an appropriate subjective workload measure [1]. Similarly, Hendy, Hamilton and Landry [5], also suggested that weighting the ratings could not add to the sensitivity of the NASA-TLX. Thus, in the present study this evaluation process was omitted. A simple arithmetic mean was computed across subscales of the NASA-TLX and it was treated as a subjective mental workload score.

The NASA-task load index was administered to all the subjects at two times i.e. at pre- main task session and post-main task session. Each subject received two mental workload rating scores on six subscales of mental workload i.e. mental demand, physical demand, temporal demand, effort, frustration and performance. The ANOVA results revealed that the main effects of mental demand ($F_{1, 76} = 16.59$; $p < .001$), temporal demand ($F_{1, 76} = 9.33$; $p < .01$), effort ($F_{1, 76} = 10.62$; $p < .01$), frustration ($F_{1, 76} = 16.59$; $p < .001$), performance ($F_{1, 76} = 73.29$; $p < .001$), and overall mental work load ($F_{1, 76} = 5.79$; $p <.01$) were significant except physical demand. These results suggested that mental demand (pre- M = 79.36; post M = 72.79), temporal demand (pre- M = 67.93; post M = 61.35), effort (pre- M = 74.08; post M = 66.91), frustration (pre- M = 32.00; post M = 21.76), and overall mental workload (pre- M = 60.99; post M = 58.75) were rated higher at pre-test than they were at their counterparts i.e. post test session, irrespective of the experimental conditions. However, subjects experienced almost equal physical demand at pre- and post-sessions (pre- M = 52.36; post M = 53.57). Moreover, rating of own performance workload (pre- M = 60.19; post M = 76.01) increased during test sessions. The performance workload is associated with the level of satisfaction the subjects felt about his/her performance in accomplishing the goal of the task. These findings supported the fourth hypothesis, which suggested that automation would reduce mental workload. To examine further the relationship between subscales of the NASA-TLX at pre- and post-test sessions, Pearson's product moment correlations were computed. The majority of correlations amongst the various measures of workload were highly significant with r-values ranging between 0.27 and 0.80. In sum, it is evident from the results that subjects experienced high mental workload on majority of the sub-scales of the NASA-TLX at the initial level i.e., pre-test, which got reduced across time periods while performing two 30-min flight simulation tasks (see Fig. 4).

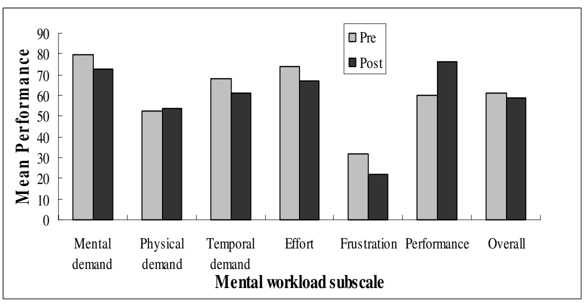

Fig. 4. Pre-and Post workload scores in NASA-TLX subscale

4 Discussion

Automation plays a critical role in situations when a small number of operators must control and supervise a very complex set of remote processes. Automation here is not optional; it is a necessity [15]. Complex machines tend to distance operators from the details of an operation. Over time, if the machines are reliable, operators will come to rely upon them, and may become less concerned with the details of the process. Though this has the desirable effect of moderating human operator workload, it also has the undesirable effect of making the operator feel less involved in the task being performed. That's where complacency comes into play within the domain of cognitive downfalls. The workload overload consequences might affect the human performance from increases in task resource demand or stressful situations. These also may have some unforeseen errors like more selectivity of input, more important sources of information given more weight, decrease in accuracy, decreasing use of strategies that involve heavy mental computation and locking into a single strategy. Most critical being the operators continuing awareness of the objective importance of all tasks that may compete for attention, for instant those of lesser importance will be shed first. Therefore the antidote may include to redesign the task by assigning some loaded tasks to automated mode and include a display design such that, information for the most tasks are available, interpretable and salient.

Several reports have discussed the dangers of automation-induced complacency. The general experience in aviation has been that advanced automated devices often do reduce workload, but usually at flight phases where workload is already low, such as cruise whereas some automation actually increases workload at critical phases, for instance at the time of take off and landing. Thus automation merely shifts the pattern of workload between work phases. However, little empirical research has been produced to substantiate its harmful effects on performance as well as some other cognitive factors that could be the root cause for automation-induced complacency. The present endeavor looks into revalidation of earlier findings of automation-induced complacency [12, 19], and also an examined the intriguing relationship among extended training reliability and mental workload. The current experimental results suggest that automation-induced complacency might be the feature of multi-task condition where subjects detected automation failures under high static system reliability condition as compared to variable system reliability condition. Imperatively, this effect of automation-induced complacency further enhances after half an hour of task period.

Considering further, the present experimental conditions also looked into the discrete relationship between subjective mental workload and automation-induced complacency. The obtained pre and post mental workload revealed that mental demand, temporal demand, effort, frustration and overall workload significantly reduced from pre to post sessions. The performance-rating workload of subjects, further significantly enhanced over time, whereas physical demand was stable during test sessions. The compilation of results suggested that high system reliability would reduce workload, thereby resulting in automation-induced complacency. And hence this finding corroborated the findings of Metzger and Parasuraman [9], who recently reported that reliable automation reduces mental workload.

5 Conclusion

Automation is rapidly being incorporated in modern flight decks thereby accounting for augmented system reliability and decreasing operational errors. However, despite it's potential for improving overall system (or aircrafts) performance, automation is not always considered good. It has been always associated with unique and unanticipated problems. Automation has enhanced the importance of the need to comprehend and control the factors that influence human monitoring behavior. Errors are least likely to show their effect or influence the results, when workload is moderate and does not alter suddenly or unpredictably [8]. Thus, this study argues for the goal to emphasize consequences of automation. Maintaining appropriate levels of workload during automated operating conditions is one of the key issues in the design of nuclear power plants and other process control environments. Therefore, over and under load continue to be a critical cognitive and human factors issue. Additional studies with different monitoring tasks, ingenious methods for adaptive control and under different scenarios of multiple task performance need to be demeanor to test effectiveness of sophisticated approaches.

References

1. Byers, J.C., Bittner, A.C., Hill, S.G.: Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In: Mital, A. (ed.) Advances in industrial ergonomics and safety I, Taylor & Francis, London (1989)
2. Comstock, J.R., Arnegard, R.J.: The multi-attribute task battery for human operator workload and strategic behavior research. (NASA Tech. Memorandum 104174). NASA Langley. Research Centre, Hampton, VA (1992)
3. Harris, W., Hancock, P.A., Arthur, E.: The effect of task load projection on automation use, performance, and workload. In: Proceedings of the 7th International Symposium on Aviation Psychology, Columbus, OH (1993)
4. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human mental workload, pp. 139–183. Human mental workload, North Holland, Amsterdam (1988)
5. Hendy, K.C., Hamilton, K.M., Landry, L.N.: Measuring subjective workload: When is one scale better than many? Human Factors 35, 579–601 (1993)
6. Hilburn, B.: Dynamic decision aiding: the impact of adaptive automation on mental workload. In: Harris, D. (ed.) Aldershot: Ashgate, pp. 193–200. Ashgate, Aldershot (1997)
7. Hughes, D., Dornheim, M.A.: Accident's direct focus on cockpit automation. Aviation Week. and Space Technology 30, 52–54 (1995)
8. Kantowitz, B.H., Casper, P.A.: Human workload in aviation. In: Wiener, E., Nagel, D. (eds.) Human factors in aviation, pp. 157–185. Academic Press, New York (1988)
9. Metzger, U., Parasuraman, R.: Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. Human Factors 47(1), 35–49 (2005)
10. Miyake, S., Kumashiro, M.: Subjective mental workload assessment technique-an introduction to NASA-TLX and SWAT and a proposal of simple scoring methods. Japanese Journal of Ergonomics 29, 399–408 (1993)
11. Parasuraman, R., Riley, V.: Humans and automation. Use, misuse, disuse, abuse. Human Factors 39, 230–253 (1997)
12. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consequences of automation-induced complacency. International Journal of Aviation Psychology 3(1), 1–23 (1993)
13. Riley, V.: A theory of operator reliance on automation. In: Mouloua, M., Parasuraman, R. (eds.) Human performance in automated systems: Current research and trends. LEA, Hillsdale, NJ, pp. 8–14 (1994)

14. Sharma, H.O., Singh, I.L.: Extended training, automation reliability and monitoring task performance. Psychological Studies, 49 (4) (2004)
15. Sheridan, T.B. (ed.): Humans and automation. System design and research issues. Wiley, New York (2002)
16. Singh, I.L., Molloy, R., Parasuraman, R.: Automation-induced complacency. Development of the complacency potential rating scale. International Journal of Aviation Psychology 3, 111–122 (1993)
17. Singh, I.L., Molloy, R., Parasuraman, R.: Automation-induced monitoring inefficiency: Role of display location. International Journal of Human Computer Human Computer Studies 46, 17–46 (1997)
18. Singh, I.L., Molloy, R., Parasuraman, R., Westerman, S.: Does location affect automation-induced complacency? In: Paper presented in the First Conference on Automation and Human Performance, Washington, DC, USA (1994)
19. Singh, I.L., Parasuraman, R.: Human performance in automated systems. Indian Psychological Abstract and Reviews 8(2), 235–276 (2001)
20. Singh, I.L., Sharma, H.O., Singh, A.L.: Effect of training on workload in flight simulation task performance. Journal of the Indian Academy of Applied Psychology 31(1-2), 81–90 (2005)
21. Wiener, E.L.: Cockpit automation. In: Wiener, E.L., Nagel, D.C. (eds.) Human factors in aviation. pp. 433–461. Academic, San Diego (1988)
22. Wiener, E.L.: Human factors of advanced technology ("glass cockpit") transport aircraft (Report No.177528). Ames Research Centre, Moffett Field, CA (1989)
23. Woods, D.D.: Automation. Apparent simplicity, real complexity. In: Mouloua, M., Parasuraman, R. (eds.) Human performance in automated systems: Current research and trends, pp. 1–7. Lawrence Erlbaum Associates, Hillsdale, NJ (1994)